

HMM filtering and parameter estimation of an electricity spot price model

Christina Erlwein ^{*} Fred Espen Benth[†] Rogemar Mamon ^{‡,*}

January 29, 2007

Abstract

In this paper we develop a model for electricity spot price dynamics. The spot price is assumed to follow an exponential Ornstein-Uhlenbeck (OU) process with an added compound Poisson process, therefore the model allows for mean-reversion and possible jumps. A sinusoidal factor is also introduced to capture the seasonality component of prices. The mean-reverting level, speed of adjustment and volatility of the OU process as well as the mean and variance of the normally distributed jump sizes of the compound Poisson process are all modulated by a hidden Markov chain in discrete time. The parameters are able to switch between different economic regimes representing various levels of supply and demand. Through the application of reference probability technique, adaptive filters are derived, which in turn, provide optimal estimates for the state of the Markov chain and related quantities of the observation process. The EM algorithm is applied to find optimal estimates of the model parameters in terms of the recursive filters. Since the parameters are updated everytime a new information is available, the model is self-calibrating. We implement the model on a deseasonalized series of daily spot electricity prices from the Nordic exchange Nord Pool. On the basis of one-step ahead forecasts, we found that the model is able to capture the stylised features of Nord Pool spot prices.

Keywords: spot electricity prices, hidden Markov model, optimal parameter estimation, EM-algorithm, Ornstein-Uhlenbeck process, jump process

^{*}CARISMA, School of Information Systems, Computing and Mathematics, Brunel University, Uxbridge, United Kingdom

[†]CMA, Centre of Mathematics for Applications, University of Oslo, Oslo, Norway

[‡]Department of Statistical and Actuarial Sciences, 2nd Floor Western Science Centre, University of Western Ontario, London, Canada

1 Introduction

Over the last two decades, electricity markets in many countries have become deregulated where spot and futures contracts are open for trade. Compared to standard financial products, electricity spot prices show distinctive stochastic properties which call for new models.

Electricity is a non-storable commodity, which leads to a strong dependency on supply and demand and therefore to high seasonal differences in prices. This seasonality shows daily, weekly and annual patterns. Additional stylised features of electricity prices include mean-reversion and frequently occurring spikes. These stochastic properties have led to different approaches for modeling electricity prices. Lucia and Schwartz [16] proposed a two-factor mean-reverting model for spot prices with a deterministic component for the seasonal pattern. Another approach was taken by Deng [8], Benth *et al.* [1] and Cartea and Figueroa [3], where the characteristics of spot prices are captured with mean-reversion dynamics driven by Lévy (jump) processes. A jump-diffusion model for hourly spot prices was proposed by Culot *et al.* [5]. The calibration and parameter estimation in these models, however, can be problematic due to limited historical data and a large number of parameters. One of the main motivations in many of these studies is to derive the futures price dynamics. Instead of modeling the spot prices for this purpose, Clewlow and Strickland [4], Benth and Koekebakker [2], and Kiesel, Schindlmayer and Börger [15] choose to model the entire forward curve directly to price forward and futures contracts.

A study by de Jong [6] found that spikes in spot electricity prices can be better captured by regime-switching models than by a Poisson jump model. Regime-switching models for electricity prices were developed by Deng [8] and De Jong and Huisman [7]. Most regime-switching models distinguish between two regimes, one ‘normal’ and one ‘jump’ regime. Huisman and Mahieu [14] introduced a third regime for the change from ‘jump’ to ‘normal’ regime. Elliott, Sick and Stein [11] introduced a Markov model to electricity spot prices. Here, the number of generators on-line are represented by a Markov process in discrete time and parameters are estimated

with the EM-algorithm. A more general version of regime-switching models are hidden Markov models (HMM). Generally, an HMM is a double-embedded stochastic process with one observation and one underlying process defined by its number of states and transition probabilities. The concept of HMM's was applied to electricity markets by Yu and Sheblé [19] describing the structure of the electricity market with an HMM, and by González, San Roque and García-González [12], who use an Input/Output HMM for analyzing electricity prices. In this paper we develop an HMM for forecasting electricity spot prices. We assume that the electricity spot price is only a partial observation. The underlying economic state, which represents the current state of supply and demand, is hidden in this observation process. The economic state is modeled by a Markov chain in discrete time, which governs the parameters of our model.

One main problem in forecasting prices on the electricity market is the estimation of parameters since daily prices can be very volatile and jumps can occur throughout the year. We develop a mean-reverting model with jumps, where the parameters evolve according to the underlying discrete time Markov chain. Following the method by Elliott [9] for estimating parameters in an HMM discrete time setting, we are able to derive recursive parameter estimates. The contribution of this paper is two-fold: the proposed model is able to capture the main features that characterize electricity spot prices (seasonality, mean-reversion and jumps) and recursive estimates for the model parameters are derived through adaptive filters for the state of the Markov chain and related processes. One step-ahead spot price forecasts are generated, which follow the actual data closely.

The paper is organized as follows: in Section 2 we describe the model framework and the underlying stochastic process. Section 3 details the derivation of the filters for the state of the Markov chain and related quantities through a change of probability measure. In section 4 these adaptive filters are used to find optimal estimates for the model parameters. These recursive formulas are derived by employing the EM algorithm and relating the parameters in the model to the processes of the Markov chain. The implementation of this model is shown in section 5. The data set consists of daily spot prices from the Nordic power exchange Nord Pool. Before the

filters are applied, the actual data set is deseasonalized. We found that a 3-state Markov chain produces one-step ahead forecasts with small mean square prediction errors. Our model therefore is capable of capturing the salient features of electricity spot price dynamics in the market. The last section presents some conclusions and remarks.

2 Model description

The spot price model for electricity is composed of two components: one deterministic function $D(k)$ to capture seasonal trends, and an Ornstein-Uhlenbeck process X with Markov-modulated parameters to model the mean-reversion of electricity prices observed in the market. The random price fluctuations are modeled by a Brownian motion W to include the ‘normal’ variations when the market is quiet and a jump process Y for the spikes. The observation process is defined on the underlying probability space (Ω, \mathcal{F}, P) . Throughout the entire discussion we denote all vectors by bold small letters and all matrices by bold capitalised letters.

Let \mathbf{z}_k be a homogeneous Markov chain with finite state in discrete time ($k = 0, 1, \dots$) and state space $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$, the canonical basis of \mathbb{R}^n . The different states of the Markov chain represent regimes with higher and lower electricity demand. Let $\mathcal{F}_k^0 = \sigma\{\mathbf{z}_0, \dots, \mathbf{z}_k\}$ be the σ -field generated by $\mathbf{z}_0, \dots, \mathbf{z}_k$, and \mathcal{F}^z be the complete filtration generated by \mathcal{F}_k^0 . Under the probability measure P the Markov chain \mathbf{z} has the dynamics $\mathbf{z}_{k+1} = \mathbf{\Pi}\mathbf{z}_k + \mathbf{v}_{k+1}$, where $\mathbf{\Pi}$ denotes the transition probability matrix of the Markov chain and \mathbf{v}_{k+1} is a martingale increment with $E[\mathbf{v}_{k+1} \mid \mathcal{F}_k] = 0$. The spot price dynamics is given by

$$S(k) = D(k) \exp(X_k) \tag{1}$$

We model the seasonal component $D(k)$ with a sinusoidal function with positive trend. The sinusoidal function includes a yearly and a weekly component, since the electricity demand shows seasonal patterns for colder and warmer times of the year

as well as for weekend- and weekday-demand. The seasonal component is given by

$$\begin{aligned} D(k) = & ak + s_1 \sin\left(\frac{2\pi}{365}k\right) + s_2 \cos\left(\frac{2\pi}{365}k\right) \\ & + s_3 \sin\left(\frac{2\pi}{7}k\right) + s_4 \cos\left(\frac{2\pi}{7}k\right) + c \end{aligned} \quad (2)$$

for some constants a , s_i and c to be determined. The stochastic processes X_t has the following dynamics

$$dX_t = \alpha(\mathbf{z}_t)(\beta(\mathbf{z}_t) - X_t)dt + \sigma(\mathbf{z}_t)dW_t + dY_t \quad (3)$$

where the level β and speed of mean-reversion α and the volatility σ are governed by the Markov chain \mathbf{z}_t . The usual Euclidean scalar product is denoted by $\langle \cdot, \cdot \rangle$ with $\alpha(\mathbf{z}_t) = \langle \boldsymbol{\alpha}, \mathbf{z}_t \rangle$, $\beta(\mathbf{z}_t) = \langle \boldsymbol{\beta}, \mathbf{z}_t \rangle$ and $\sigma(\mathbf{z}_t) = \langle \boldsymbol{\sigma}, \mathbf{z}_t \rangle$. The jump process Y_t is given by

$$dY_t = JdN_t, \quad (4)$$

where N_t is a Poisson process with constant intensity λ and jump sizes J_i . The jump sizes J_i are dependent on the Markov chain \mathbf{z}_t meaning that different demand regimes have different jump size distributions. The conditional distribution of the jump sizes is $J_i|z_t \sim N(\mu_J(z_t), \sigma_J^2(z_t))$. The intensity λ does not change when a switching of regimes occurs. The seasonality of jump intensity is still taken into account, since the jump size is evolving according to the state of the Markov chain. The filtration generated by the observation process is defined by $\mathcal{F}^X = \sigma(X_1, X_2, \dots)$ and includes the filtration generated by the Brownian motion \mathcal{F}^W and the filtration generated by the jump process component \mathcal{F}^Y . The global filtration is defined by $\mathcal{F} = \mathcal{F}^X \vee \mathcal{F}^Z$.

3 Filtering

In this section we derive adaptive filters for processes of the Markov chain \mathbf{z} . We use a change of measure technique, so we are able to derive recursive filters under a new ideal measure, where the calculations are easier.

3.1 Change of measure

We define our observation process as the logarithm of the deseasonalized electricity spot prices, which is therefore given by

$$\begin{aligned} X_t &= \ln \frac{S(t)}{D(t)} \\ &= X_s e^{-\alpha(\mathbf{z}_s)(t-s)} + \beta(\mathbf{z}_s)(1 - e^{-\alpha(\mathbf{z}_s)(t-s)}) \\ &\quad + \sigma(\mathbf{z}_s) e^{-\alpha(\mathbf{z}_s)t} \int_s^t e^{\alpha(\mathbf{z}_s)u} dW_u + \sum_{m=N_s+1}^{N_t} e^{-\alpha(\mathbf{z}_s)(t-\tau_m)} J_m(\mathbf{z}_t). \end{aligned} \quad (5)$$

The parameters α, β, σ and the jump-size J_t of the compound Poisson process component are governed by a Markov chain \mathbf{z}_t in discrete time. The random time of occurrence of the m -th jump is denoted by τ_m . For deriving filters of related processes of the Markov chain \mathbf{z} and for finding optimal parameter estimates we work under a reference probability measure \bar{P} . To do this we need a discrete version of our observation process. Discretizing equation (5) leads to

$$\begin{aligned} X_{k+1} &= X_k e^{-\alpha(\mathbf{z}_k)\Delta k} + \beta(\mathbf{z}_k)(1 - e^{-\alpha(\mathbf{z}_k)\Delta k}) + \sigma(\mathbf{z}_k) \sqrt{\frac{1 - e^{-2\alpha(\mathbf{z}_k)\Delta k}}{2\alpha(\mathbf{z}_k)}} h_{k+1} \\ &\quad + \sum_{m=1}^{N_{\Delta k}} e^{-\alpha(\mathbf{z}_k)(\Delta k - \tau_m)} J_m(\mathbf{z}_k) \end{aligned} \quad (6)$$

where \mathbf{z}_k is a discrete time Markov chain and $\{h_{k+1}\}$ is a sequence of IID standard normal random variables. Note the following connection for the discretization of the jump-term:

$$\int_l^k e^{-\alpha(\mathbf{z}_k)(k-u)} dY_u = \sum_{m=N_k+1}^{N_{k+1}} e^{-\alpha(\mathbf{z}_k)(k-\tau_m)} J_m(\mathbf{z}_k) \underbrace{=}_{\text{in distr.}} e^{-\alpha(\mathbf{z}_k)(k-l)} \sum_{m=1}^{N_{k+1}-N_k} e^{\alpha(\mathbf{z}_k)\tau_m} J_m(\mathbf{z}_k),$$

where τ_m are the jumping times in the interval $(0, k-l]$.

We calculate our filters under a reference probability measure \bar{P} . Under this measure \mathbf{z} is still a Markov chain with dynamics $\mathbf{z}_{k+1} = \mathbf{\Pi} \mathbf{z}_k + \mathbf{v}_{k+1}$ and X_k are independent observations. To perform a change of measure we examine the discretized observation process. Note that we assume that the change of measure does not affect

the compound Poisson process component of the observation process. As in Merton [17] the jump size and intensity have the same dynamics under the new measure.

We construct a reference probability measure \bar{P} by applying a discrete time version of Girsanov's theorem. Setting the Radon-Nikodym derivative to $\frac{d\bar{P}}{dP} |_{\mathcal{F}_k} = \Lambda_k = \prod_{l=1}^k \lambda_l$ we define

$$\begin{aligned} \lambda_l : &= \exp \left[\frac{1}{d} \left[X_l e^{-\alpha(\mathbf{z}_l)\Delta l} + \beta(\mathbf{z}_l)(1 - e^{-\alpha(\mathbf{z}_l)\Delta l}) + \sum_{m=1}^{N_{\Delta l}} e^{-\alpha(\mathbf{z}_l)(\Delta l - \tau_m)} J_m(\mathbf{z}_l) \right] h_{l+1} \right. \\ &\quad \left. - \frac{1}{2d^2} \left[X_l e^{-\alpha(\mathbf{z}_l)\Delta l} + \beta(\mathbf{z}_l)(1 - e^{-\alpha(\mathbf{z}_l)\Delta l}) + \sum_{m=1}^{N_{\Delta l}} e^{-\alpha(\mathbf{z}_l)(\Delta l - \tau_m)} J_m(\mathbf{z}_l) \right]^2 \right] \quad (7) \end{aligned}$$

with $d = \sigma(\mathbf{z}_l) \sqrt{\frac{1 - e^{-2\alpha(\mathbf{z}_l)\Delta l}}{2\alpha(\mathbf{z}_l)}}$. The process $\{\Lambda_l\}$ is a P -almost surely positive martingale with filtration \mathcal{F} , $E^P[\Lambda] = 1$. Under P , \mathbf{z} is a Markov chain with the same transition matrix $\mathbf{\Pi}$. Following Elliott, Aggoun and Moore (1995) we back out the real world measure from the reference probability measure by defining $\frac{dP}{d\bar{P}} |_{\mathcal{F}_k} = \bar{\Lambda}_k = \prod_{l=1}^k \bar{\lambda}_l$ with

$$\begin{aligned} \bar{\lambda}_l : &= \exp \left[-\frac{1}{d} \left[X_l e^{-\alpha(\mathbf{z}_l)\Delta l} + \beta(\mathbf{z}_l)(1 - e^{-\alpha(\mathbf{z}_l)\Delta l}) + \sum_{m=1}^{N_{\Delta l}} e^{-\alpha(\mathbf{z}_l)(\Delta l - \tau_m)} J_m(\mathbf{z}_l) \right] \frac{X_{l+1}}{d} \right. \\ &\quad \left. - \frac{1}{2d^2} \left[X_l e^{-\alpha(\mathbf{z}_l)\Delta l} + \beta(\mathbf{z}_l)(1 - e^{-\alpha(\mathbf{z}_l)\Delta l}) + \sum_{m=1}^{N_{\Delta l}} e^{-\alpha(\mathbf{z}_l)\tau_m} J_m(\mathbf{z}_l) \right]^2 \right] \quad (8) \end{aligned}$$

where $\bar{\Lambda}_0 = 1$, $\{\bar{\lambda}_l : \bar{\lambda} \in \mathbb{N}^+\}$ and $\{\bar{\Lambda}_l : l \in \mathbb{N}\}$. The process $\{\bar{\Lambda}_l\}$ is a \mathcal{F} -martingale under P and $\bar{\Lambda}\Lambda = 1$.

Therefore we found an equivalent probability measure which is used in the following to calculate adaptive filters for Markov chain processes. With Bayes theorem, a filter for any adapted process H is given by

$$E[H_k | \mathcal{F}_k^X] = \frac{\bar{E}[H_k \bar{\Lambda}_k | \mathcal{F}_k^X]}{\bar{E}[\bar{\Lambda}_k | \mathcal{F}_k^X]}.$$

Write $\eta(H_k) := \bar{E}[H_k \bar{\Lambda}_k | \mathcal{F}_k^X]$, so that $E[H_k | \mathcal{F}_k^X] = \frac{\eta(H_k)}{\eta(1)}$. We can derive recursive filters for the term $\eta(H_{k-1} \mathbf{z}_{k-1})$. This conditional expectation is related to

the desired term $\eta(H_{k-1})$ through $\langle \mathbf{1}, \eta(H_k \mathbf{z}_k) \rangle = \eta(H_k)$. Therefore

$$E[H_k | \mathcal{F}_k^X] = \frac{\langle \mathbf{1}, \eta(H_k \mathbf{z}_k) \rangle}{\langle \mathbf{1}, \eta(\mathbf{z}_k) \rangle}.$$

We derive filters for the state space process of the Markov chain, the jump process G , the occupation time process O and for auxiliary processes T . These filters are calculated according to Elliott (1994). To calculate recursive estimates we define Γ^i according to the new measure

$$\begin{aligned} \Gamma^i = & \exp \left[-\frac{1}{d_i^2} [X_l e^{-\alpha_i \Delta l} + \beta_i (1 - e^{-\alpha_i \Delta l}) + \sum_{m=1}^{N_{\Delta l}} e^{-\alpha_i (\Delta l - \tau_m)} J_m^i] X_{l+1} \right. \\ & \left. - \frac{1}{2d_i^2} [X_l e^{-\alpha_i \Delta l} + \beta_i (1 - e^{-\alpha_i \Delta l}) + \sum_{m=1}^{N_{\Delta l}} e^{-\alpha_i (\Delta l - \tau_m)} J_m^i]^2 \right]. \end{aligned} \quad (9)$$

Following Elliott (1994) we find recursive filters for the Markov chain processes. The state estimator is then given by

$$\eta_k(\mathbf{z}_k) = \sum_{i=1}^N \Gamma^i(\mathbf{z}_k) \langle \mathbf{e}_i, \eta_{k-1}(\mathbf{z}_{k-1}) \rangle \mathbf{\Pi e}_i. \quad (10)$$

The jump process G from state r to state s of the Markov process, defined as $G_k^{(sr)} = \sum_{l=1}^k \langle \mathbf{z}_{l-1}, \mathbf{e}_r \rangle \langle \mathbf{z}_l, \mathbf{e}_s \rangle$ has the recursive filter

$$\begin{aligned} \eta_k(G_k^{sr} \mathbf{z}_k) = & \sum_{i=1}^N \Gamma^i(\mathbf{z}_k) \langle \eta_{k-1}(G_{k-1}^{sr} \mathbf{z}_{k-1}), \mathbf{e}_i \rangle \mathbf{\Pi e}_i \\ & + \Gamma^r(\mathbf{z}_k) \eta_{k-1}(\langle \mathbf{z}_{k-1}, \mathbf{e}_r \rangle) \pi_{sr} \mathbf{e}_s. \end{aligned} \quad (11)$$

Now we calculate a filter for the occupation time process $O_k^r = \sum_{l=1}^k \langle \mathbf{z}_{l-1}, \mathbf{e}_r \rangle$, denoting the occupation time of the process at state r up to time k . We get

$$\begin{aligned} \eta_k(O_k^r \mathbf{z}_k) = & \sum_{i=1}^N \Gamma^i(\mathbf{z}_k) \langle \eta_{k-1}(O_{k-1}^r \mathbf{z}_{k-1}), \mathbf{e}_i \rangle \mathbf{\Pi e}_i \\ & + \Gamma^r(\mathbf{z}_k) \langle \eta_{k-1}(\mathbf{z}_{k-1}), \mathbf{e}_r \rangle \mathbf{\Pi e}_r. \end{aligned} \quad (12)$$

For the calculation of the optimal parameter estimates we need an auxiliary process T , which is given by $T_k^{(r)}(f) = \sum_{l=1}^k \langle \mathbf{z}_{l-1}, \mathbf{e}_r \rangle f(X_l)$ where f is a function of the form

$f(X) = X_l$, $f(X) = X_l^2$, $f(X) = X_{l+1}X_l$ or $f(X) = X_{l+1}^2$, $1 \leq l \leq k$. The recursive filter for this process is

$$\begin{aligned} \eta_k(T_k^r(f)\mathbf{z}_k) &= \sum_{i=1}^N \Gamma^i(\mathbf{z}_k) \{ \langle \eta_{k-1}(T_{k-1}^r(f)\mathbf{z}_{k-1}), \mathbf{e}_i \rangle \mathbf{\Pi e}_i \\ &\quad + \Gamma^r(\mathbf{z}_k) \langle \eta_{k-1}(\mathbf{z}_{k-1}), \mathbf{e}_r \rangle f(X_k) \mathbf{\Pi e}_r \}. \end{aligned} \quad (13)$$

4 Optimal parameter estimates

In this section we want to derive a maximum likelihood estimation for the parameters of the observation process X_t (see equation 5), and a jump diffusion process where the parameters are governed by a Markov chain \mathbf{z}_t .

First we derive the probability density function (pdf) for the process X_t

$$\begin{aligned} X_t &= X_s e^{-\alpha(\mathbf{z}_s)(t-s)} + \beta(\mathbf{z}_s)(1 - e^{-\alpha(\mathbf{z}_s)(t-s)}) \\ &\quad + \sigma(\mathbf{z}_s) e^{-\alpha(\mathbf{z}_s)t} \int_s^t e^{\alpha(\mathbf{z}_s)u} dW_u + \sum_{m=N_s+1}^{N_t} e^{-\alpha(\mathbf{z}_s)(t-\tau_m)} J_m(\mathbf{z}_t). \end{aligned} \quad (14)$$

The parameters are said to be constant over every interval $[s, t]$, $0 \leq s \leq t$. The observation process without jumps is normally distributed with mean $\mu_x = \beta + (X_s - \beta)e^{-\alpha(t-s)}$ and variance $\sigma_x^2 = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha(t-s)})$. Now we examine the distribution of the part given by the compound Poisson process Y_t . As described in the previous section J_1, J_2, \dots are independent, identically distributed normal random variables and $(N_t)_{t \geq 0}$ is a standard Poisson process with jump intensity $\lambda > 0$. Let N and J be jointly independent. We denote the mean and the variance of the process J by μ_J and σ_J^2 respectively. The probability distribution of the Poisson process N is given by the usual Poisson distribution. To derive the density of the jump component we can do the following approximation of the jump integral

$$\int_s^t e^{-\alpha(t-u)} dY_u \approx e^{-\alpha(t-s)}(Y_t - Y_s).$$

By the stationarity of the compound Poisson process, we find that the increment $Y_t - Y_s$ has the same distribution as Y_{t-s} , and thus we have the following density of

the contribution from the jump term

$$\Phi_{Y_{t-s}}(x) = \sum_{h=0}^{\infty} \frac{(\lambda(t-s))^h}{h!} e^{-\lambda(t-s)} \phi(x; \mu_J e^{-\alpha(t-s)} h, \sigma_J^2 e^{-2\alpha(t-s)} h) \quad (15)$$

where ϕ denotes the pdf of the normal distribution.

Following the arguments by Hanson and Westman (2001) the pdf of our observation process can be calculated as the convolution of densities of the OU process without jumps and the jump part distribution. We therefore have the density of X_t conditioned on X_s as

$$\begin{aligned} \Phi_X(x) &= \sum_{h=0}^{\infty} \frac{(\lambda(t-s))^h}{h!} e^{-\lambda(t-s)} \phi\left(x; \beta + (X_s - \beta)e^{-\alpha(t-s)} + \mu_J e^{-\alpha(t-s)} h, \right. \\ &\quad \left. \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha(t-s)}) + \sigma_J^2 e^{-2\alpha(t-s)} h\right). \end{aligned} \quad (16)$$

The density in equation (16) can be further expressed as an expectation of the normal density under the Poisson counter $N_{\Delta t}$. The density (16) can be written as

$$\begin{aligned} \Phi_X(x) &= E_{N_{\Delta k}} \left[\phi(x; \beta + (X_s - \beta)e^{-\alpha(t-s)} + \mu_J e^{-\alpha(t-s)} N_{\Delta t}, \right. \\ &\quad \left. \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha(t-s)}) + \sigma_J^2 e^{-2\alpha(t-s)} N_{\Delta t} \right]. \end{aligned} \quad (17)$$

We wish to find the optimal parameters of the observation process X_t specified in equation (5) using the EM algorithm. For this purpose we make the simplifying assumption, that the intensity of the Poisson process λ is independent from the other parameters. To find optimal estimates, we evaluate the parameters of the normal distributed part of the observation process independent of the process N_t . Therefore we first derive the maximum likelihood estimates (MLE) for the set of parameters $\xi = \{\alpha_i, \beta_i, \sigma_i^2, \mu_{J_i}, \sigma_{J_i}^2, \pi_{ji}\}$. Our aim is to find a new set of parameters $\hat{\xi}$, which maximises the conditional expectation of the log-likelihoods. In the following we denote the jump counter $N_{\Delta k}$ with p and the mean and variance of the OU-process with μ_x and σ_x respectively. We derive MLE's for the normal distribution $\tilde{\phi}(x; \mu_x + \mu_J p e^{-\alpha(t-s)}, \sigma_x^2 + \sigma_J^2 p e^{-\alpha(t-s)})$. We note that both mean and variance are dependent on the Markov chain \mathbf{z} , they are therefore regime-switching. The discretized version of the observation process (see equation 6) is used for deriving

the recursive parameter updates.

We derive an explicit recursive formula for the parameter β with the processes of the Markov chain \mathbf{z} . However, since the mean-reversion level α is included in the mean and variance part, the calculation of the MLE for α is less straightforward and a recursive formula cannot be found. We therefore derive an explicit recursive formula for the mean μ_x and the mean of the jump process μ_J and calculate a value for α based on the optimal value of β by solving the equation $\mu_{x_i} = \beta_i + (X_l - \beta_i)e^{-\alpha_i\Delta}$. Consequently

$$\alpha_i = -\ln\left(\frac{\mu_{x_i} - \beta_i}{X_l - \beta_i}\right)\frac{1}{\Delta} . \quad (18)$$

With the value of α from (18) together with the MLE estimate of σ_x the estimated value of σ_i is given by

$$\sigma_i^2 = \frac{2\alpha_i\sigma_{x_i}}{1 - e^{-2\alpha_i\Delta}} . \quad (19)$$

Therefore, calculating MLE's for μ_x and σ_x gives us the desired parameter estimates for α and σ^2 .

Applying the EM algorithm we derive the following optimal recursive parameter estimates:

$$\hat{\mu}_{x_i} = \frac{\tilde{T}_k^i(X_{k+1}) - \tilde{O}_k^i\mu_{J_i}pe^{-\alpha_i\Delta}}{\tilde{O}_k^i} \quad (20)$$

$$\hat{\mu}_{J_i} = \frac{\tilde{T}_k^i(X_{k+1}) - \tilde{O}_k^i\mu_{x_i}}{\tilde{O}_k^i pe^{-\alpha_i\Delta}} \quad (21)$$

$$\begin{aligned} \hat{\beta}_i &= \frac{\tilde{T}_k^i(X_l)(e^{-2\alpha_i\Delta} + e^{-\alpha_i\Delta}) + \tilde{T}_k^i(X_{l+1})(1 - e^{-\alpha_i\Delta})}{\tilde{O}_k^i(1 + 2e^{-\alpha_i\Delta} + e^{-2\alpha_i\Delta})} \\ &\quad - \frac{\tilde{O}_k^i(-e^{-2\alpha_i\Delta}\mu_{J_i}p + \mu_{J_i}pe^{-\alpha_i\Delta})}{\tilde{O}_k^i(1 + 2e^{-\alpha_i\Delta} + e^{-2\alpha_i\Delta})} \end{aligned} \quad (22)$$

$$\begin{aligned}\hat{\sigma}_{x_i}^2 &= \frac{\tilde{T}_k^i(X_{k+1}^2) + \tilde{O}_k^i(\mu_{x_i}^2 + \mu_{J_i}^2 p^2 e^{-2\alpha_i \Delta} + 2\mu_{x_i} \mu_{J_i} p e^{-\alpha_i \Delta} - \sigma_{J_i}^2 e^{-2\alpha_i \Delta} p)}{\tilde{O}_k^i} \\ &\quad - \frac{2\tilde{T}_k^i(X_{k+1})(\mu_{x_i} + \mu_{J_i} p e^{-\alpha_i \Delta})}{\tilde{O}_k^i}\end{aligned}\quad (23)$$

$$\begin{aligned}\hat{\sigma}_{J_i}^2 &= \frac{\tilde{T}_k^i(X_{k+1}^2) + \tilde{O}_k^i(\mu_{x_i}^2 + \mu_{J_i}^2 p^2 e^{-2\alpha_i \Delta} + 2\mu_{x_i} \mu_{J_i} p e^{-\alpha_i \Delta} - \sigma_{x_i}^2)}{\tilde{O}_k^i p e^{-2\alpha_i \Delta}} \\ &\quad - \frac{2\tilde{T}_k^i(X_{k+1})(\mu_{x_i} + \mu_{J_i} p e^{-\alpha_i \Delta})}{\tilde{O}_k^i p e^{-2\alpha_i \Delta}}\end{aligned}\quad (24)$$

and

$$\hat{\pi}_{ji} = \frac{\tilde{G}_k^{ji}}{\tilde{O}_k^i}. \quad (25)$$

The proofs for equation (20) – (25) can be found in the Appendix.

5 Implementation

The model is implemented on daily spot prices compiled by Nordpool. The data set SP contains daily spot prices from 1998-2002.

5.1 Fitting the deterministic function

First we deseasonalize the data. The deterministic function is fitted to the actual data. The parameters for the deterministic function are calibrated with a least-square algorithm in Matlab. In particular, $\frac{1}{2} \sum_t (D(x, k) - SP(k))^2$ is minimized with respect to x , where x denotes a set of parameters. Here $x = \{a, s_1, s_2, s_3, s_4, c\}$, it includes all parameters from equation (2). The resulting deterministic function for the seasonal components is

$$\begin{aligned}D(k) &= 0.0569k + 14.1033 \sin\left(\frac{2\pi}{365}k\right) + 20.6332 \cos\left(\frac{2\pi}{365}k\right) \\ &\quad + 8.5458 \sin\left(\frac{2\pi}{7}k\right) - 0.5251 \cos\left(\frac{2\pi}{7}k\right) + 97.2454\end{aligned}\quad (26)$$

In Figure 1 the spot prices in NOK/MWh are depicted together with the seasonal function. Frequent jumps in the electricity prices are visible, the descriptive statistics show a high variance of the price data.

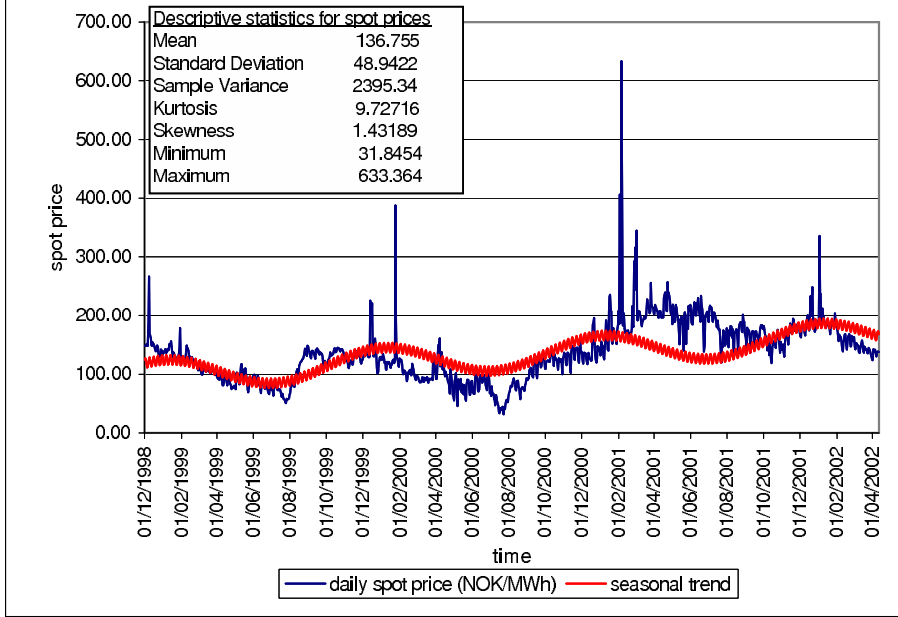


Figure 1: **Actual data and seasonal function**

The remaining stochastic part is the log of the deseasonalized spot price S . We consider this as our observation process for the empirical work presented in the next subsection.

5.2 Filtering and parameter estimation

The filters for updating the parameters are applied to the data set. We calculate a series of one-step ahead forecasts for the spot prices. The expected value of the observation process at time $k + 1$ is calculated with

$$\begin{aligned}
 E[X_{k+1} \mid \mathcal{G}_k] &= X_k e^{-\langle \alpha, \Pi \mathbf{z}_k \rangle \Delta k} + \langle \beta, \Pi \mathbf{z}_k \rangle (1 - e^{-\langle \alpha, \Pi \mathbf{z}_k \rangle \Delta k}) \\
 &\quad + \lambda k e^{-\langle \alpha, \Pi \mathbf{z}_k \rangle \Delta k} \langle \mu_J, \Pi \mathbf{z}_k \rangle
 \end{aligned} \tag{27}$$

We use 1200 data points between December 1998 and March 2002 to obtain one-step ahead forecasts. The model parameters are updated after processing the data in batches of 60 data points. The algorithm is run twenty times within this data set. The implementation is performed under the set-up of a 2-state and 3-state Markov chain. The Markov chain represents therefore either a ‘high’ and ‘low’ or a ‘high’, ‘medium’ and ‘low’ economic state of electricity supply and demand. Figure 2 depicts the dynamic movement of the optimal parameter estimates. The parameters α and σ are calculated through the updated optimal parameters β, μ_x and σ_x ; all other parameters are calculated via the recursive filter estimates.

In Figure 3 the evolution of parameters in a 3-state markov chain setting is displayed. Here, the evolution of parameters exhibits similar pattern to that of the two-state Markov chain setting. The convergence of parameter estimates is slightly faster than that in the 2-state set-up.

The one-step ahead forecasts for electricity spot prices in a 3-state HMM is depicted in Figure 4. Here we can see that the one-step ahead forecast follows the actual values very closely. The self-tuning algorithm is able to capture the dynamics of the electricity spot prices and the occurrence of jumps is picked up by the filter. In one time period the forecasted values are slightly overestimating the actual values. This might be due to the fact, that after the occurrence of a significant jump downwards, the filters for the parameters first have to receive new information in order to allow for a close one-step ahead forecast once again. A comparison of the 2- and 3-state forecast shows, that the mean-square error in a 2-state setting is 0.0457. The mean-square error between the actual data and the one-step ahead forecast in a 3-state setting is 0.0421 and therefore slightly lower than in the two-state setting. It is apparent that the 3-state HMM yields better fit than the 2-state HMM. We extended the implementation to a 4-state HMM but no significant further improvement was found.

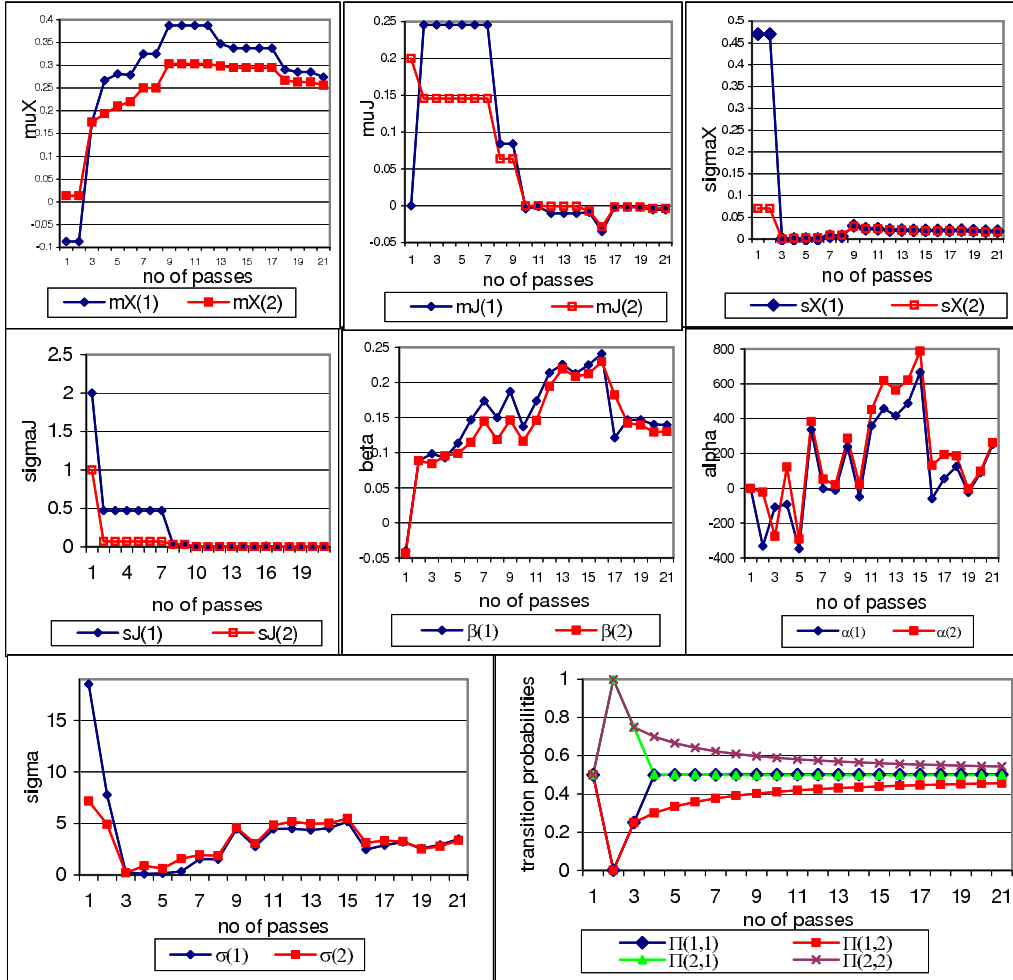


Figure 2: Parameter evolution in a 2-state HMM

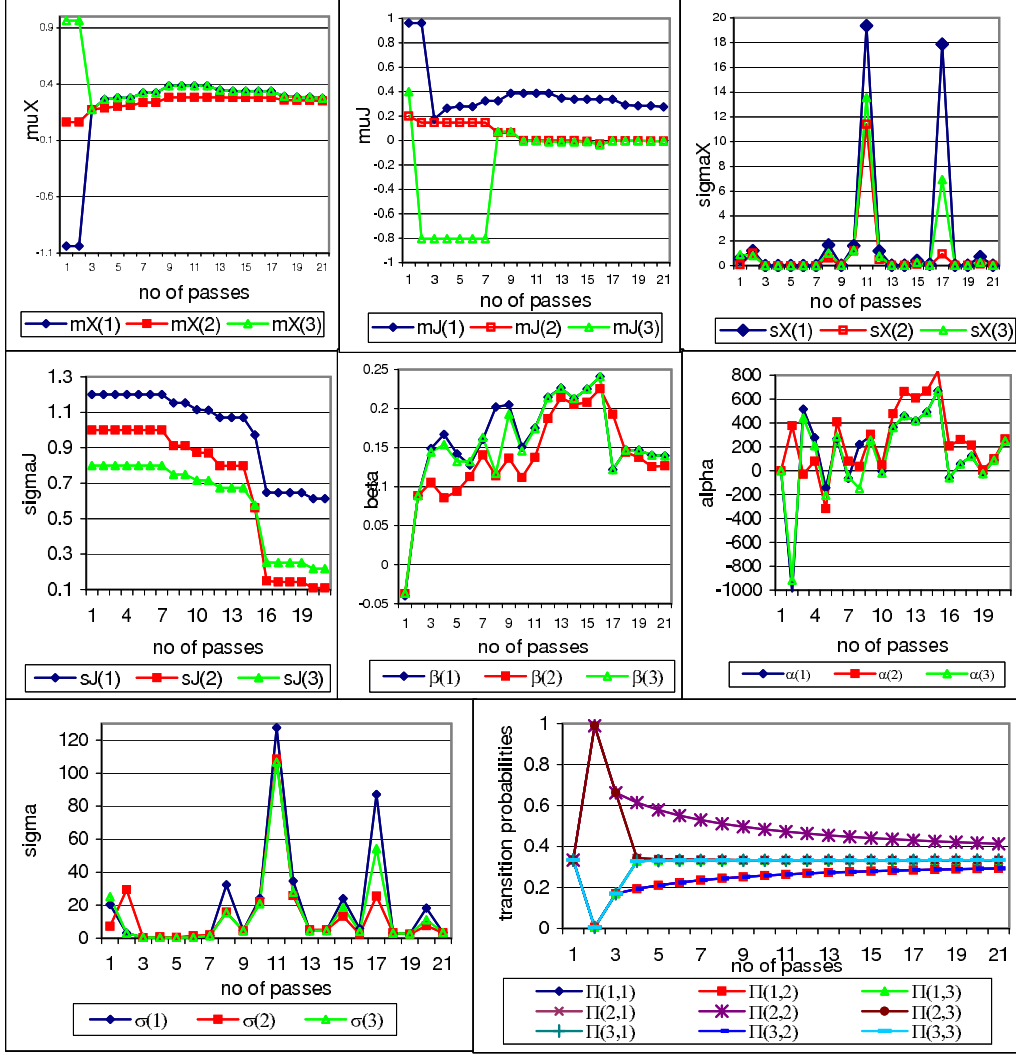


Figure 3: Parameter evolution in a 3-state HMM

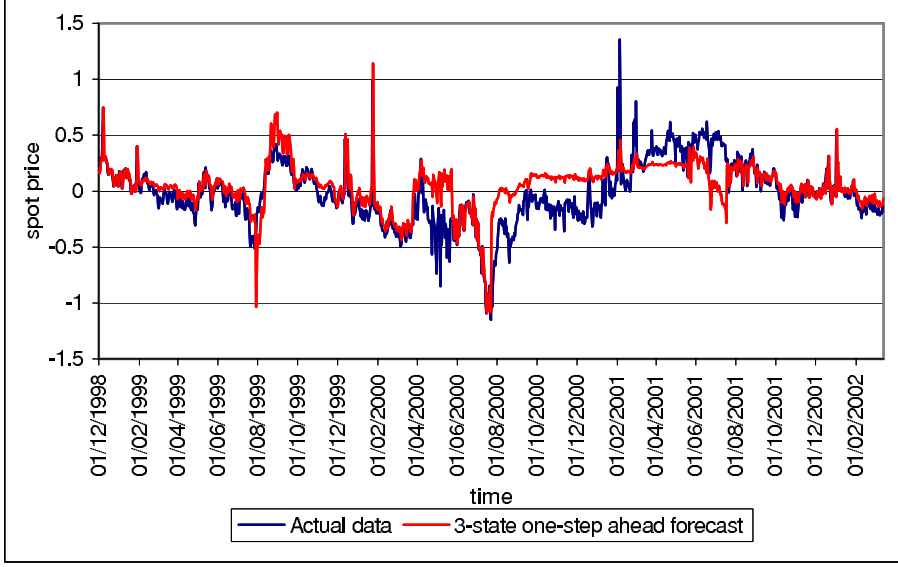


Figure 4: **One-step ahead forecasts of electricity spot prices**

6 Concluding remarks

We developed an HMM-driven model to forecast electricity spot prices. The spot price is assumed to evolve in accordance with the exponential of an OU process plus a jump term and this exponential is scaled by a deterministic sinusoidal function to take into account the seasonal component of electricity prices. The added compound Poisson process has normally distributed jumps, where the mean and variance are governed by a discrete-time HMM. This offers the model greater flexibility to switch between economic regimes reflected by the dynamic changes in electricity supply and demand, which is easily seen in the sudden jumps of spot prices. Employing the EM algorithm, the optimal estimates for the model parameters are derived in terms of the recursive filters for the state of the Markov chain, the number of jumps between two states, occupation time of the Markov chain and an auxiliary process. Since the parameters are updated whenever a new dataset arrives, we have created a self-tuning model. The empirical work on the implementation of filters and parameter estimation of the model using deseasonalized electricity spot prices illustrates that the proposed model is well-equipped to capture the spikes present in the data for both the 2-state and 3-state setting. The important stylized characteristics of the

electricity markets are captured by the model as evidenced by low forecast errors and similar trends portrayed by the forecasts relative to dynamics of the actual data series.

7 Acknowledgements

The authors would like to thank Steen Koekebakker for helpful discussions and for providing electricity spot price data for the empirical analysis. Christina Erlwein wishes to thank the financial support provided by a Marie Curie Fellowship for Early Stage Researchers Training, and also to acknowledge the hospitality of the Centre of Mathematics for Applications, University of Oslo where this work was completed.

A Appendix

A.1 Optimal parameter estimate for μ_x

We define a new measure \hat{P} by

$$\left. \frac{d\hat{P}}{dP} \right|_{\mathcal{F}_k} = \Lambda_k^* = \prod_{l=1}^k \lambda_l^*$$

where

$$\begin{aligned} \lambda_l^* &= \frac{\exp\left[-\frac{1}{2\sigma_x^2}[X_{l+1} - \hat{\mu}_x - \mu_J p e^{-\alpha\Delta}]^2\right]}{\exp\left[-\frac{1}{2\sigma_x^2}[X_{l+1} - \mu_x - \mu_J p e^{-\alpha\Delta}]^2\right]} \\ &= \exp\left[\frac{1}{2\sigma_x^2}(-(X_{l+1} - \hat{\mu}_x - \mu_J p e^{-\alpha\Delta})^2 + (X_{l+1} - \mu_x - \mu_J p e^{-\alpha\Delta})^2)\right]. \end{aligned} \quad (28)$$

The log-likelihood for Λ_k^* is

$$\begin{aligned} \log \Lambda_k^* &= \sum_{l=1}^k \left[-\frac{1}{2\sigma_x^2}(-\hat{\mu}_x^2 + 2X_{l+1}\hat{\mu}_x - 2\hat{\mu}_x\mu_J p e^{-\alpha\Delta} + \mu_x^2 \right. \\ &\quad \left. - 2X_{l+1}\mu_x + 2\mu_x\mu_J p e^{-\alpha\Delta}) \right]. \end{aligned} \quad (29)$$

We substitute the processes of the Markov chain \mathbf{z} into this log-likelihood and get

$$\log \Lambda_k = \sum_{i=1}^n \left[-\frac{1}{2\sigma_{x_i}^2 O_k^i} [\hat{\mu}_{x_i}^2 O_k^i - 2T_k^i(X_{k+1})\hat{\mu}_{x_i} + 2O_k^i \hat{\mu}_{x_i} \mu_{J_i} p e^{-\alpha_i \Delta} + R(\mu_x)] \right] \quad (30)$$

where $R(\mu_x)$ is a remainder without $\hat{\mu}$. Now, the conditional expectation of the log-likelihood $L(\hat{\mu}_{X_i}) = E[\log \Lambda_k | \mathcal{F}_k^X]$ is considered. For any process H write $\tilde{H}_l = E[H_l | \mathcal{F}_k^X]$.

$$L(\hat{\mu}_{x_i}) = \sum_{i=1}^n \left[-\frac{1}{2\sigma_{x_i}^2 \tilde{O}_k^i} [\tilde{O}_k^i \mu_{x_i}^2 - 2\tilde{T}_k^i(X_{k+1})\hat{\mu}_{x_i} + 2\tilde{O}_k^i \hat{\mu}_{x_i} \mu_{J_i} p e^{-\alpha_i \Delta} + R(\mu_x)] \right] \quad (31)$$

We differentiate $L(\hat{\mu}_{x_i})$ in $\hat{\mu}_{x_i}$ and equate the result to 0. That gives

$$2\tilde{O}_k^i \hat{\mu}_{x_i} - 2\tilde{T}_k^i(X_{k+1}) + 2\tilde{O}_k^i \mu_{J_i} p e^{-\alpha_i \Delta} = 0 \quad (32)$$

$$\text{or} \quad \hat{\mu}_{x_i} = \frac{\tilde{T}_k^i(X_{k+1}) - \tilde{O}_k^i \mu_{J_i} p e^{-\alpha_i \Delta}}{\tilde{O}_k^i}. \quad (33)$$

A.2 Optimal parameter estimate for μ_J

The new measure is defined by the Radon-Nikodym derivative as above with

$$\lambda_l^* = \frac{\exp\left[-\frac{1}{2\sigma_x^2}(X_{l+1} - \mu_x - \hat{\mu}_J p e^{-\alpha \Delta})^2\right]}{\exp\left[-\frac{1}{2\sigma_x^2}(X_{l+1} - \mu_x - \mu_J p e^{-\alpha \Delta})^2\right]} \quad (34)$$

$$= \exp\left[-\frac{1}{2\sigma_x^2}(\hat{\mu}_J^2 p^2 e^{-2\alpha \Delta} - 2X_{l+1}\hat{\mu}_J p e^{-\alpha \Delta} + 2\mu_x \hat{\mu}_J p e^{-\alpha \Delta} + R(\mu_J))\right] \quad (35)$$

where $R(\mu_J)$ is a remainder without $\hat{\mu}_J$. We calculate the log-likelihood and include the processes of the Markov chain \mathbf{z} :

$$\log \Lambda_k^* = \sum_{i=1}^n \left[-\frac{1}{2\sigma_{x_i}^2 O_k^i} [O_k^i \hat{\mu}_{J_i}^2 p^2 e^{-2\alpha_i \Delta} - 2T_k^i(X_{k+1})\hat{\mu}_{J_i} p e^{-\alpha_i \Delta} + 2O_k^i \mu_{x_i} \hat{\mu}_{J_i} p e^{-\alpha_i \Delta} + R(\mu_J)] \right]. \quad (36)$$

Write $\tilde{H}_l = E[H_l | \mathcal{F}_k^X]$. Now, differentiate the log-likelihood in $\hat{\mu}_{J_i}$ and equate the differential to 0

$$2\tilde{O}_k^i \hat{\mu}_{J_i} p^2 e^{-2\alpha_i \Delta} - 2\tilde{T}_k^i(X_{k+1}) p e^{-\alpha_i \Delta} + 2\tilde{O}_k^i \mu_{x_i} p e^{-\alpha_i \Delta} = 0 \quad (37)$$

$$\text{or} \quad \hat{\mu}_{J_i} = \frac{\tilde{T}_k^i(X_{k+1}) - \tilde{O}_k^i \mu_{x_i}}{\tilde{O}_k^i p e^{-\alpha_i \Delta}}. \quad (38)$$

A.3 Optimal parameter estimate for β

We define the new measure \hat{P} by

$$\left. \frac{d\hat{P}}{dP} \right|_{\mathcal{F}_k} = \Lambda_k^* = \prod_{l=1}^k \lambda_l^* \quad (39)$$

where

$$\begin{aligned} \lambda_l^* &= \frac{\exp\left[-\frac{1}{2\sigma_x^2} [X_{l+1} - \hat{\beta} - [X_l - \hat{\beta}]e^{-\alpha\Delta} - \mu_J p e^{-\alpha\Delta}]^2\right]}{\exp\left[-\frac{1}{2\sigma_x^2} [X_{l+1} - \beta - [X_l - \beta]e^{-\alpha\Delta} - \mu_J p e^{-\alpha\Delta}]^2\right]} \\ &= \exp\left[-\frac{1}{2\sigma_x^2} \left[(X_{l+1}^2 + \hat{\beta}^2 + ([X_l - \hat{\beta}]e^{-\alpha\Delta})^2 + (\mu_J p e^{-\alpha\Delta})^2 - 2X_{l+1}\hat{\beta} \right. \right. \\ &\quad - 2X_{l+1}[X_l - \hat{\beta}]e^{-\alpha\Delta} - 2X_{l+1}\mu_J p e^{-\alpha\Delta} - 2\hat{\beta}[X_l - \hat{\beta}]e^{-\alpha\Delta} \\ &\quad + 2\hat{\beta}\mu_J p e^{-\alpha\Delta} + 2[X_l - \hat{\beta}]e^{-\alpha\Delta}\mu_J p e^{-\alpha\Delta}) - (X_{l+1}^2 + \beta^2 + ([X_l - \beta]e^{-\alpha\Delta})^2 \\ &\quad + (\mu_J p e^{-\alpha\Delta})^2 - 2X_{l+1}\beta - 2X_{l+1}[X_l - \beta]e^{-\alpha\Delta} - 2X_{l+1}\mu_J p e^{-\alpha\Delta} \\ &\quad \left. \left. - 2\beta[X_l - \beta]e^{-\alpha\Delta} + 2\beta\mu_J p e^{-\alpha\Delta} + 2[X_l - \beta]e^{-\alpha\Delta}\mu_J p e^{-\alpha\Delta}) \right] \right]. \quad (40) \end{aligned}$$

For the log-likelihood we have

$$\begin{aligned} \log \Lambda_k^* &= \sum_{l=1}^k \left[-\frac{1}{2\sigma_x^2} \left[\hat{\beta}^2(1 + e^{-2\alpha\Delta} + 2e^{-\alpha\Delta}) + 2\hat{\beta}(-e^{-2\alpha\Delta}X_l - X_{l+1} \right. \right. \\ &\quad \left. \left. + X_{l+1}e^{-\alpha\Delta} - X_l e^{-\alpha\Delta} + \mu_J p e^{-\alpha\Delta} - e^{-2\alpha\Delta}\mu_J p) + R(\beta) \right] \right] \quad (41) \end{aligned}$$

where $R(\beta)$ is a remainder which does not include $\hat{\beta}$ terms. We substitute the expressions including the Markov chain with the defined processes $O_k^i = \sum_{l=1}^k \langle \mathbf{z}_l, \mathbf{e}_i \rangle$ and $T_k^i(f) = \sum_{l=1}^k \langle \mathbf{z}_l, \mathbf{e}_i \rangle f(X_l)$. Therefore

$$\begin{aligned} \log \Lambda_k^* &= \sum_{i=1}^n \left[-\frac{1}{2\sigma_{x_i}^2} O_k^i \left[\hat{\beta}_i^2 (O_k^i(1 + e^{-2\alpha_i\Delta} + 2e^{-\alpha_i\Delta}) + 2\hat{\beta}_i(-T_k^i(X_l)e^{-2\alpha_i\Delta} \right. \right. \\ &\quad - T_k^i(X_{l+1}) + T_k^i(X_{l+1})e^{-\alpha_i\Delta} - T_k^i(X_l)e^{-\alpha\Delta} \\ &\quad \left. \left. + O_k^i(-e^{-2\alpha_i\Delta}\mu_{Ji}p + \mu_{Ji}p e^{-\alpha_i\Delta})) \right] + R(\beta) \right]. \quad (42) \end{aligned}$$

To calculate the expectation of the log-likelihood conditional on \mathcal{F}_k^X we set $\tilde{H}_l = E[H_l \mid \mathcal{F}_k^X]$ for any process H . We differentiate $L(\hat{\beta})$ in $\hat{\beta}_i$ and equate the result to 0. This gives

$$2\hat{\beta}_i \tilde{O}_k^i (1 + 2e^{-\alpha_i \Delta} + e^{-2\alpha_i \Delta}) - 2(\tilde{T}_k^i(X_l)(e^{-2\alpha_i \Delta} + e^{-\alpha_i \Delta}) - 2\tilde{T}_k^i(X_{l+1})(1 - e^{-\alpha_i \Delta}) + 2\tilde{O}_k^i(-e^{-2\alpha_i \Delta} \mu_{J_i} p + \mu_{J_i} p e^{-\alpha_i \Delta})) = 0.$$

Henceforth

$$\begin{aligned} \hat{\beta}_i &= \frac{\tilde{T}_k^i(X_l)(e^{-2\alpha_i \Delta} + e^{-\alpha_i \Delta}) + \tilde{T}_k^i(X_{l+1})(1 - e^{-\alpha_i \Delta})}{\tilde{O}_k^i(1 + 2e^{-\alpha_i \Delta} + e^{-2\alpha_i \Delta})} \\ &\quad - \frac{\tilde{O}_k^i(-e^{-2\alpha_i \Delta} \mu_{J_i} p + \mu_{J_i} p e^{-\alpha_i \Delta})}{\tilde{O}_k^i(1 + 2e^{-\alpha_i \Delta} + e^{-2\alpha_i \Delta})}. \end{aligned} \quad (43)$$

A.4 Optimal parameter estimate for σ_x

For the MLE of the variance of the Ornstein-Uhlenbeck component in the observation process, we define the Radon-Nikodym derivative $\frac{d\hat{P}}{dP}$ with following λ^*

$$\begin{aligned} \lambda_l^* &= \frac{\frac{1}{\sqrt{\hat{\sigma}_x^2 + \sigma_J^2 e^{-2\alpha \Delta} p}} \exp\left[\frac{1}{2(\hat{\sigma}_x^2 + \sigma_J^2 e^{-2\alpha \Delta} p)}(X_{l+1} - \mu_x - \mu_J p e^{-\alpha \Delta})^2\right]}{\frac{1}{\sqrt{\sigma_x^2 + \sigma_J^2 e^{-2\alpha \Delta} p}} \exp\left[\frac{1}{2(\sigma_x^2 + \sigma_J^2 e^{-2\alpha \Delta} p)}(X_{l+1} - \mu_x - \mu_J p e^{-\alpha \Delta})^2\right]} \\ &= \frac{\sqrt{\sigma_x^2 + \sigma_J^2 e^{-2\alpha \Delta} p}}{\sqrt{\hat{\sigma}_x^2 + \sigma_J^2 e^{-2\alpha \Delta} p}} \exp\left[-\frac{1}{2(\hat{\sigma}_x^2 + \sigma_J^2 e^{-2\alpha \Delta} p)}(X_{l+1} - \mu_x - \mu_J p e^{-\alpha \Delta})^2\right] \\ &\quad + \frac{1}{2(\sigma_x^2 + \sigma_J^2 e^{-2\alpha \Delta} p)}(X_{l+1} - \mu_x - \mu_J p e^{-\alpha \Delta})^2. \end{aligned}$$

The log-likelihood of Λ_k is therefore

$$\begin{aligned} \log \Lambda_k^* &= \sum_{k=1}^l \left(\frac{1}{2} \log \frac{\sigma_x^2 + \sigma_J^2 e^{-2\alpha \Delta} p}{\hat{\sigma}_x^2 + \sigma_J^2 e^{-2\alpha \Delta} p} - \frac{1}{2(\hat{\sigma}_x^2 + \sigma_J^2 e^{-2\alpha \Delta} p)}(X_{l+1} - \mu_x - \mu_J p e^{-\alpha \Delta})^2 \right. \\ &\quad \left. + R(\sigma_x^2) \right). \end{aligned} \quad (44)$$

Since σ_x is regime-switching, we have the following conditional expectation of the log-likelihood including the Markov chain \mathbf{z} :

$$\begin{aligned}
L(\hat{\sigma}_x^2) &= \sum_{l=1}^k \sum_{i=1}^n \left(-\frac{1}{2} \langle \mathbf{z}_l, \mathbf{e}_i \rangle \log(\hat{\sigma}_{x_i}^2 + \sigma_{J_i}^2 e^{-2\alpha_i \Delta} p) \right. \\
&\quad \left. - \langle \mathbf{z}_l, \mathbf{e}_i \rangle \frac{1}{2(\sigma_{x_i}^2 + \sigma_{J_i}^2 e^{-2\alpha_i \Delta} p)} (X_{l+1} - \mu_{x_i} - \mu_{J_i} p e^{-\alpha_i \Delta})^2 \right) + R(\sigma_x^2) \\
&= \sum_{i=1}^n \left(-\frac{1}{2} \tilde{O}_k^i \log(\hat{\sigma}_{x_i}^2 + \sigma_{J_i}^2 e^{-2\alpha_i \Delta} p) \right. \\
&\quad - \frac{1}{2(\hat{\sigma}_{x_i}^2 + \sigma_{J_i}^2 e^{-2\alpha_i \Delta} p)} (\tilde{T}_k^i(X_{k+1}^2) + \tilde{O}_k^i \mu_{x_i}^2 + \tilde{O}_k^i \mu_{J_i}^2 p^2 e^{-2\alpha_i \Delta} - 2\tilde{T}_k^i(X_{k+1}) \mu_{x_i} \\
&\quad \left. - 2\tilde{T}_k^i(X_{k+1}) \mu_{J_i} p e^{-\alpha_i \Delta} + 2\tilde{O}_k^i \mu_{x_i} \mu_{J_i} p e^{-\alpha_i \Delta}) \right) + R(\sigma_x^2). \tag{45}
\end{aligned}$$

To find the maximum we differentiate $L(\hat{\sigma}_x^2)$ in each $\hat{\sigma}_{x_i}^2$ and equate the resulting derivative to 0.

$$\begin{aligned}
&\tilde{T}_k^i(X_{k+1}^2) + \tilde{O}_k^i(\mu_{x_i}^2 + \mu_{J_i}^2 p^2 e^{-2\alpha_i \Delta} + 2\mu_{x_i} \mu_{J_i} p e^{-\alpha_i \Delta}) - 2\tilde{T}_k^i(X_{k+1})(\mu_{x_i} + \mu_{J_i} p e^{-\alpha_i \Delta}) \\
&= \tilde{O}_k^i(\hat{\sigma}_{x_i}^2 + \sigma_{J_i}^2 e^{-2\alpha_i \Delta} p). \tag{46}
\end{aligned}$$

We find the following optimal parameter estimate for $\hat{\sigma}_{x_i}^2$

$$\begin{aligned}
\hat{\sigma}_{x_i}^2 &= \frac{\tilde{T}_k^i(X_{k+1}^2) + \tilde{O}_k^i(\mu_{x_i}^2 + \mu_{J_i}^2 p^2 e^{-2\alpha_i \Delta} + 2\mu_{x_i} \mu_{J_i} p e^{-\alpha_i \Delta} - \sigma_{J_i}^2 p e^{-2\alpha_i \Delta})}{\tilde{O}_k^i} \\
&\quad - \frac{2\tilde{T}_k^i(X_{k+1})(\mu_{x_i} + \mu_{J_i} p e^{-\alpha_i \Delta})}{\tilde{O}_k^i}. \tag{47}
\end{aligned}$$

A.5 Optimal parameter estimate for σ_J^2

The Radon-Nikodym derivative $\frac{d\hat{P}}{dP}$ is defined with

$$\begin{aligned}
\lambda_l^* &= \sqrt{\frac{\sigma_x^2 + \sigma_J^2 p e^{-2\alpha \Delta}}{\sigma_x^2 + \hat{\sigma}_J^2 e^{-2\alpha \Delta} p}} \exp \left[-\frac{1}{2(\sigma_x^2 + \hat{\sigma}_J^2 e^{-2\alpha \Delta} p)} (X_{l+1} - \mu_x - \mu_J p e^{-\alpha \Delta})^2 \right. \\
&\quad \left. + \frac{1}{2(\sigma_x^2 + \hat{\sigma}_J^2 e^{-2\alpha \Delta} p)} (X_{l+1} - \mu_x - \mu_J p e^{-\alpha \Delta}) \right]. \tag{48}
\end{aligned}$$

Therefore the log-likelihood of Λ_k^* is given by

$$\begin{aligned} \log \Lambda_k^* &= \sum_{l=1}^k \left(-\frac{1}{2} \log(\sigma_x^2 + \hat{\sigma}_J^2 e^{-2\alpha\Delta} p) \right. \\ &\quad \left. - \frac{1}{2(\sigma_x^2 + \hat{\sigma}_J^2 e^{-2\alpha\Delta} p)} (X_{l+1} - \mu_x - \mu_J p e^{-\alpha\Delta})^2 + R(\sigma_J^2 p) \right). \end{aligned} \quad (49)$$

With that the conditional expectation is

$$\begin{aligned} L(\hat{\sigma}_J^2) &= \sum_{i=1}^n \left(-\frac{1}{2} \tilde{O}_k^i \log(\sigma_{x_i}^2 + \hat{\sigma}_{J_i}^2 e^{-2\alpha_i\Delta} p) - \frac{1}{2(\sigma_{x_i}^2 + \hat{\sigma}_{J_i}^2 e^{-2\alpha_i\Delta} p)} (\tilde{T}_k^i(X_{k+1}^2) \right. \\ &\quad \left. + \tilde{O}_k^i(\mu_{x_i}^2 + \mu_{J_i}^2 p^2 e^{-2\alpha_i\Delta} + 2\mu_{x_i}\mu_{J_i} p e^{-\alpha_i\Delta}) \right. \\ &\quad \left. - 2\tilde{T}_k^i(X_{k+1})(\mu_{x_i} + \mu_{J_i} p e^{-\alpha_i\Delta}) \right) + R(\sigma_J^2 p) \end{aligned} \quad (50)$$

Differentiating (50) with respect to $\hat{\sigma}_{J_i}^2$ gives

$$\begin{aligned} \hat{\sigma}_{J_i}^2 &= \frac{\tilde{T}_k^i(X_{k+1}^2) + \tilde{O}_k^i(\mu_{x_i}^2 + \mu_{J_i}^2 p^2 e^{-2\alpha_i\Delta} + 2\mu_{x_i}\mu_{J_i} p e^{-\alpha_i\Delta} - \sigma_{x_i}^2)}{\tilde{O}_k^i p e^{-2\alpha_i\Delta}} \\ &\quad - \frac{2\tilde{T}_k^i(X_{k+1})(\mu_{x_i} + \mu_{J_i} p e^{-\alpha_i\Delta})}{\tilde{O}_k^i p e^{-2\alpha_i\Delta}} \end{aligned} \quad (51)$$

A.6 Optimal parameter estimate for the transition probabilities π_{ij}

The optimal estimate for the transition probabilities is calculated by considering the Radon-Nikodym derivative

$$\begin{aligned} \left. \frac{d\hat{P}}{dP} \right|_{\mathcal{F}_k} &= \Lambda_k^* = \prod_{l=1}^k \left(\sum_{r,s=1}^n \left(\frac{\hat{\pi}_{sr}}{\pi_{sr}} \right)^{\langle X_l, \mathbf{e}_s \rangle \langle X_{l-1}, \mathbf{e}_r \rangle} \right) \\ \text{with } \Lambda_0^* &= 1. \end{aligned}$$

We obtain the log-likelihood

$$\log \Lambda_k^* = \sum_{s,r}^n J_k^{sr} \log \hat{\pi}_{sr} + R(\pi)$$

where the remainder does not involve $\hat{\pi}_{sr}$. We maximise the log-likelihood subject to the constraint $\sum_i^n \hat{\pi}_{sr} = 1$. Consequently, the optimal estimates for the parameter Π is

$$\hat{\pi}_{sr} = \frac{\tilde{G}_k^{ji}}{\tilde{O}_k^i}.$$

References

- [1] Benth, F.E., Ekeland, L., Hauge, R., Nielsen, B., 2003. A note on arbitrage-free pricing of forward contracts in energy markets. *Applied Mathematical Finance* 10 (4), 325-336.
- [2] Benth, F.E. and Koekebakker, S., 2005. Stochastic modeling of financial electricity contracts. E-print no. 24, Department of Mathematics, University of Oslo.
- [3] Cartea, A. and Figueroa, M., 2005. Pricing in electricity markets: a mean reverting jump diffusion model with seasonality. *Applies Mathematical Finance* 12 (4), 313-335.
- [4] Clewlow, L. and Strickland, C., 2000. Valuing Energy Options in a one factor model to forward prices. Working paper, University of Technology, Sydney.
- [5] Culot, M., Goffin, V., Lawford, S., de Menten, S., 2006. An affine jump diffusion model for electricity. Working paper, Université Catholique de Louvain.
- [6] De Jong, C., 2005. The nature of power spikes: a regime-switch approach. ERIM Report Series Reference No. ERS-2005-052-F&A Available at SSRN: <http://ssrn.com/abstract=828173>
- [7] De Jong, C. and Huisman, R., 2002. Option Formulas for Mean-Reverting Power Prices with Spikes. Energy Global Research Paper; ERIM Report Series Reference No. ERS-2002-96-F&A. Available at SSRN: <http://ssrn.com/abstract=324520>

- [8] Deng, S., 1998. Stochastic models of energy commodity prices and their applications: mean-reversion with jumps and spikes. Technical report, Georgia Institute of Technology.
- [9] Elliott, R., 1994. Exact adaptive filters for Markov chains observed in Gaussian noise. *Automatica* 30, 1399-1408.
- [10] Elliott, R., Aggoun, L., Moore, J., 1995. Hidden Markov models. Springer, New York.
- [11] Elliott, R., Sick, G., Stein, M., 2003. Modelling electricity price risk. Technical Report, University of Calgary.
- [12] González, A.M., Roque, A.M.S., Garcia-Gonzalez, J., 2005. Modeling and forecasting electricity prices with input/output hidden Markov models. *IEEE Transactions on Power Systems* 20, 13-24.
- [13] Hanson, F., Westman, J., 2002. Stochastic Analysis of Jump-Diffusions for Financial Log-Return processes. In: *Proceedings of Stochastic Theory and Control Workshop*, pp 169-184. *Lecture Notes in Control and Information Sciences*, B.Pasik-Duncan (Editor), Springer-Verlag, New York.
- [14] Huisman, R. and Mahieu, R., 2003. Regime jumps in electricity prices. *Energy Economics* 25, 425-434.
- [15] Kiesel, R., Schindlmayr, G., and Börger, R. (2006). A two-factor model for the electricity forward market. *Preprint*, University of Ulm, Germany.
- [16] Lucia, J. and Schwartz, E.S. 2002. Electricity prices and power derivatives: evidence from the Nordic Power Exchange. *Review of Derivatives Research* 5, 5-50.
- [17] Merton, R. C., 1976. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* 3, 125-144.
- [18] Shiryaev, A.N., 1999. *Essentials of Stochastic Finance: Facts, Models, Theory*. World Scientific Publishing, Singapore.

- [19] Yu, W. and Sheblé, G., 2005. Modeling electricity markets with hidden Markov model. *Electric Power Systems Research* 76, 445-451.